Genetic analyses reveal high levels of seed and pollen flow in hawthorn (Crataegus monogyna Jacq.), a key component of hedgerows

Brown, James A.; Beatty, Gemma; Finlay, Caroline M. V.; Montgomery, W. Ian; Tosh, David G.; Provan, James

Published in:
Tree Genetics and Genomes

DOI:
10.1007/s11295-016-1020-0

Publication date:
2016

Citation for published version (APA):
Brown, J. A., Beatty, G., Finlay, C. M. V., Montgomery, W. I., Tosh, D. G., & Provan, J. (2016). Genetic analyses reveal high levels of seed and pollen flow in hawthorn (Crataegus monogyna Jacq.), a key component of hedgerows. *Tree Genetics and Genomes*, *12*(3), [58]. https://doi.org/10.1007/s11295-016-1020-0

ORIGINAL PAPER

Genetic analyses reveal high levels of seed and pollen flow in hawthorn (*Crataegus monogyna* Jacq.), a key component of hedgerows

James A. Brown[1] · Gemma E. Beatty[1,2] · Caroline M. V. Finlay[1] · W. Ian Montgomery[1,2,3] · David G. Tosh[1,2] · Jim Provan[1,2,4*]

[1] School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL

[2] *Quercus*, School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL

[3] Institute for Global Food Security, Queen's University Belfast

[4] Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Penglais, Aberystwyth SY23 3DA

* For correspondence: Dr.Jim Provan
IBERS, Aberystwyth University
Tel: +44 1970 62 2324
E-mail: J.Provan@aber.ac.uk

Abstract Hedgerows represent important components of agri-environment landscapes that are increasingly coming under threat from climate change, emergent diseases, invasive species and land-use change. Given that population genetic data can be used to inform best-practice management strategies for woodland and hedgerow tree species, we carried out a study on hawthorn (*Crataegus monogyna* Jacq.), a key component of hedgerows, on a regional basis using a combination of nuclear and chloroplast microsatellite markers. We found that levels of genetic diversity were high and comparable to, or slightly higher than, other tree species from the same region. Levels of population differentiation for both sets of markers, however, were extremely low, suggesting extensive gene flow via both seed and pollen. These findings suggest that a holistic approach to woodland management, one which does not necessarily rely on the concept of "seed zones" previously suggested, but which also takes into account populations with high and/or rare chloroplast (i.e. seed-specific) genetic variation, might be the best approach to restocking and replanting.

KEYWORDS**:** *Crataegus monogyna*, gene flow, genetic diversity, hawthorn, microsatellites, woodland management

Introduction

Hedgerows represent an important component of agri-environment landscapes, not just as boundaries for fields, but also in providing a wide range of ecosystem services such as habitats, food sources and wildlife corridors for animals, acting as wind-breaks, and preventing soil erosion (reviewed in Burel and Baudry 1995). The last few decades have seen the emergence of an increasing number of threats to tree and shrub species, including those which make up hedgerows. These threats include plant pathogens and pests, habitat loss, land use change, invasive species and climate change (Rackham 2008). Consequently, managing hedgerow species, especially trees, is becoming increasingly important, requiring clear knowledge and understanding of their ecological attributes and requirements. Trees lost due to disease or other factors may have to be replaced under management programmes. In Great Britain, the Forestry Commission has drawn up a map of 24 areas (seed zones) with the goal of maintaining provenance by only restocking woodlands with seed from the same zone (Herbert et al. 1999). These zones are defined by similar geographic, climatic and ecological characteristics, but it is becoming clear that population genetic structuring can also play an important role in woodland management (Müller-Starck et al. 1992; Ennos et al. 1998). Recent studies on ash (*Fraxinus excelsior* L.; Sutherland et al. 2010; Beatty et al. 2015a) and alder (*Alnus glutinosa* [L.] Gaertn.; Beatty et al. 2015b) found very little genetic differentiation between woodlands across Great Britain and Ireland, suggesting that the sourcing of local seeds for replanting might not be required.

Natural recolonization, regeneration and succession in plants, including trees, relies on seed dispersal, which ultimately determines adult vegetation composition (Howe and Smallwood 1982; Nathan and Muller-Landau 2000; Levine and Murrell 2003). Likewise, pollen dispersal also has an important role to play in the same processes, especially across

deforested landscapes (reviewed in Bacles and Ennos 2008, but see Provan et al. 2008). Measuring seed dispersal is difficult, with both "tagging" and "trapping" methods having drawbacks (reviewed in Beatty et al. 2015a). The use of high-resolution chloroplast microsatellites markers, however, permits the elucidation of maternal gene flow (i.e. via seed in angiosperms; Provan et al. 2001), and in combination with data from nuclear markers can indicate the relative importance of seed and pollen flow in natural populations (Ennos 1994). Furthermore, fine-scale studies wherein both parent plants and saplings are sampled allow individual parental assignment and the ability to directly measure seed dispersal distances, as demonstrated in a recent study on European ash (Beatty et al. 2015a).

Hawthorn (*Crataegus monogyna*, also known as common or one-seeded hawthorn) is a species which grows in both woodland and hedgerows, generally occurring primarily in the latter (Gosler 1990). It prefers areas with a high light intensity and exposed soil, and is adept at colonising abandoned or eroded areas (Fineschi et al. 2005). Hawthorn hedges provide habitats for many vertebrate and invertebrate species, and are important nesting sites for birds, with their thorns providing protection from predators (Pollard et al 1974; Fineschi et al. 2005). The species is indigenous to Europe, and is also found from North Africa to the Himalayas (Christensen 1992). Like all hawthorns, *C. monogyna* facilitates more than one group of pollinators, including primarily bumblebees, honey bees and hoverflies (Gyan and Woodell 1987; Gosler 1990; Campbell et al. 1991). The fruits it produces persist throughout the winter, acting as an important food source for birds, which are the major vectors for seed dispersal. These small, fleshy fruits each contain a single seed, and are produced from open pollination, with self-pollination generally prevented by gametophytic self-incompatibility (Clapham et al. 1990). However, a study on a British population showed the production of fruits in the absence of insect pollination, indicating either selfing or apomixis (reviewed in Jacobs et al. 2009).

In recent years, hawthorns have increasingly come under threat from several diseases, including fire blight, which is caused by the bacterial pathogen *Erwinia amylovora* (Burrill) Winslow *et al.* (Schroth et al. 1974), and leaf spot, which results from a fungal infection by *Entomosporium maculatum* Lev. (Stowell and Backus 1966). Given that very little is known on the genetic structure of natural hawthorn populations, and that population genetic studies on hedgerow species are rare compared to those on woodland trees, we analysed the genetic diversity in populations of hawthorn on a regional basis across Northern Ireland to inform management strategies. A combination of nuclear and chloroplast microsatellites was used to determine the relative importance of pollen- and seed-mediated gene flow, and to facilitate comparison with woodland species across the same region (Beatty et al. 2015a, 2015b, 2016).

Materials and methods

Sampling and DNA extraction

Samples were collected from 23 sites across Northern Ireland along with one site in the Republic of Ireland (Figure 1; Table 1). Samples were taken from a combination of hedgerows and woodland, depending on the site. Woodlands were selected that had been previously designated as ancient or semi-natural based on data collected for the Woodland Trust Inventory of ancient and long-established woodland in Northern Ireland (www.backonthemap.org.uk) and the National Survey of Native Woodlands 2003–08 in the Republic of Ireland (www.npws.ie). Woodlands were also selected based on government information from the Department of the Environment such ASSIs (Areas of Special Scientific Interest), as well as the landscape character areas listing the woodlands and species present in each region (https://www.doeni.gov.uk). During sampling, 3-4 leaves were taken from each of a maximum of 30 trees and stored in silica gel. The GPS coordinates of each tree were recorded. DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1987). Nuclear genotypes were obtained for between 19 and 30 individuals per population (Table 1; $N = 677$; mean = 28.208), and chloroplast haplotypes were obtained for between 20 and 30 individuals per population (Table 1; $N = 701$; mean = 29.208).

Genotyping

All samples were genotyped for eight nuclear microsatellites (Table 2) and six chloroplast microsatellites (Table 3). Seven of the nuclear microsatellite loci (excluding CH02D11) had previously been used in a population genetic study of two different *Crataegus* species, *C. douglasii* and *C. suksdorfii* (Lo et al. 2009). All eight microsatellite markers were originally

developed for studies of apple (*Malus* x *domestica*; Liebhard et al. 2002). To develop *de novo* chloroplast microsatellite markers, *Crataegus monogyna* chloroplast sequences in the GenBank database were searched for mononucleotide repeats of nine or more (Provan et al. 2001). Primers were designed using the Primer3 program to amplify the six loci in two multiplexes, which were combined for a single genotyping run (Table 3).

PCR was carried out in a total volume of 10 μl containing 100 ng genomic DNA, 5 pmol of 6-FAM labelled M13 primer, 0.05 pmol of each M13-tailed forward primer, 5 pmol each reverse primer, 1× PCR reaction buffer, 200 μM each dNTP, 2.5 mM $MgCl_2$ and 0.25 U GoTaq Flexi DNA polymerase (Promega, Sunnyvale, CA, USA). PCR was carried out on a number of machines: the MWG Primus thermal cycler (Ebersberg, Germany), MJ Research PTC-200 and PTC-220 Gradient Peltier thermal cyclers (Quebec, Canada) and Biometra T-Gradient thermal cycler (Göttingen, Germany) using the following conditions: initial denaturation at 94 °C for 3 min followed by a range of cycles – between 30 and 55 for the nuclear loci (see Table 2) and 30 for the chloroplast loci – of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, extension at 72 °C for 30 s and a final extension at 72 °C for 5 min. Genotyping was carried out on an AB3730xl capillary genotyping system. (Applied Biosystems, Foster City, CA, USA). Allele sizes were scored using the GENEMARKER software (V1.8, Softgenetics).

Data analysis

GENEPOP (V3.4; Raymond and Rousset, 1995) was used to test for linkage disequilibrium between nuclear microsatellite loci. To estimate genetic diversity within the populations, levels of observed (H_O) and expected (H_E) heterozygosity were calculated using the ARLEQUIN software package (V3.5.1.2; Excoffier and Lischer, 2010), whilst levels of allelic richness (A_R) and fixation indices (F_{IS}) were calculated using the FSTAT software package

(V2.9.3.2; Goudet, 2001). Significance of F_{IS} was determined by 10,000 randomisation steps. Chloroplast microsatellite allele sizes were combined into haplotypes, and levels of genetic diversity (H) based on haplotype frequencies were calculated using ARLEQUIN.

The overall level of genetic differentiation between populations was estimated using Φ_{ST}, which gives an analogue of F_{ST} (Weir and Cockerham, 1984) calculated within the analysis of molecular variance (AMOVA) framework (Excoffier et al. 1992) using ARLEQUIN. To further identify possible patterns of genetic structuring, the software package BAPS (V6; Corander et al. 2003) was used to identify clusters of genetically similar populations using a Bayesian approach. Ten replicates were run for all possible values of the maximum number of clusters (K) up to K = 24, the number of populations sampled, with a burn-in period of 10,000 iterations followed by 100,000 iterations. Multiple independent runs always gave the same outcome.

It has been shown previously that the chloroplast genome is maternally inherited in all Rosaceae tested to date (Fineschi 2005). Thus, the pollen to seed migration ratio (r) was calculated using the formula:

$$r = \frac{mp}{ms} = \frac{\left(\frac{1}{\Phi STn} - 1\right)(1+FIS) - 2\left(\frac{1}{\Phi STc} - 1\right)}{\left(\frac{1}{\Phi STc} - 1\right)} \quad \text{(Ennos 1994)}$$

where m_p and m_s are the pollen and seed migration rates respectively, F_{IS} is the mean inbreeding coefficient over all loci, and Φ_{STn} and Φ_{STc} are the levels of genetic differentiation calculated for nuclear and chloroplast markers respectively.

Results

No significant evidence of consistent linkage disequilibrium (i.e. involving the same loci) was detected between any of the eight nuclear microsatellites analysed (41 out of 694 tests). Between twelve (CH05G11) and 36 (CH01F02) alleles were detected per locus, with a total of 206 (mean = 25.75 per locus; Table 2). Levels of observed (H_O) and expected (H_E) heterozygosity ranged from 0.600 (CH06G11) to 0.924 (CH01F02; mean = 0.764), and from 0.568 (CH05G11) to 0.931 (CH05D04; mean = 0.803), respectively. Levels of F_{IS} ranged from -0.033 (CH05G11) to 0.130 (CH03C02), with a mean value of 0.047.

Within populations, levels of allelic richness (A_R) averaged over loci ranged from 8.779 (Glenarm) to 12.074 (Portaferry) with a mean value of 10.556 (Table 1). Levels of observed (H_O) and expected (H_E) heterozygosity ranged from 0.684 (Belleek) to 0.836 (Keady; mean = 0.763), and from 0.769 (Belleek) to 0.838 (Portaferry; mean = 0.803) respectively. Levels of F_{IS} ranged from -0.022 (Keady) to 0.122 (Belleek) with 13 out of the 24 values being significantly different from zero. No significant difference was observed for any of the above diversity measures between hedgerow and woodland populations.

Five of the six chloroplast microsatellite loci studied were polymorphic in the samples analysed, exhibiting between two and seven alleles (Table 3). Combining allele sizes across loci gave 23 haplotypes, twelve of which were found in a single individual (Table 1). The most common (H1) was found in 75% (523 out of 701) of the trees studied, and the four most common haplotypes (H1-H4) accounted for 93% of individuals. Levels of haplotype diversity (H) ranged from 0.193 (Drumshanbo Lough and Belleek) to 0.713 (Eglinton; Table 1).

Levels of population differentiation based on the nuclear and chloroplast markers calculated from the AMOVA were $\Phi_{ST(n)} = 0.0092$ and $\Phi_{ST(c)} = 0.0385$ respectively. The

171 BAPS analysis assigned all 24 populations to a single genetic cluster, including the Coolure
172 population from the Republic of Ireland. Finally, the ratio of pollen to seed migration (*r*) was
173 2.515.

Discussion

The results of the present study, the first to apply high-resolution microsatellite markers to examine patterns of genetic diversity in *Crataegus monogyna*, indicate high levels of diversity, but low levels of population differentiation, suggesting extensive gene flow via both seed and pollen. Previous studies on hawthorn have either used low-resolution (chloroplast restriction fragment length polymorphism – cpRFLP) markers over large geographic scales to study the phylogeography of the species, thus focusing more on long-term historical factors (Fineschi et al. 2005), or have used nuclear markers exhibiting relatively limited variation (randomly amplified polymorphic DNA – RAPD) at more local scales (Ferrazzini et al. 2008). The combination of nuclear and chloroplast microsatellites allows the identification of any fine-scale structuring of genetic variation, as well as providing insights into the relative contributions of seed and pollen flow in natural populations (Powell et al. 1996; Provan et al. 2001).

The lack of comparable studies (i.e. using similar markers over similar geographical scales) means that it is somewhat difficult to put the levels of diversity observed in the present study into context. The sole study examining nuclear genetic variation in *C. monogyna* used dominant RAPD markers (Ferrazzini et al. 2008), and thus it is not surprising that the reported value of expected heterozygosity (mean H_E = 0.291) is much lower than that in this study (mean H_E = 0.803). Likewise, the only other application of nuclear microsatellite markers to *Crataegus* species was carried out in the polyploid species *C. suksdorfii* and *C. douglasii*, meaning that comparable diversity statistics (e.g. H_E) could not be calculated (Lo et al. 2009). With regard to other tree species from the same region analysed using microsatellites, levels of diversity in hawthorn were comparable to those reported in ash (mean H_E = 0.765; Beatty et al. 2015a), but slightly higher than those reported

for alder (mean H_E = 0.663; Beatty et al. 2015b), sessile oak (mean H_E = 0.720; Beatty et al. 2016) and pedunculate oak (mean H_E = 0.714; Beatty et al. 2016). Levels of inbreeding, measured as F_{IS}, were similar to those in ash (Beatty et al. 2015a), but lower than those reported for alder (Beatty 2015b), probably as a result of the more patchy distribution of the latter in Northern Ireland.

The only previous study using chloroplast markers in *C. monogyna* found four haplotypes across the whole of Europe using cpRFLP (Fineschi et al. 2005). As in the present study, one of these haplotypes was found in the majority of individuals studied, including those from Great Britain (no samples from Ireland were analyzed) but it seems likely that this corresponds to the dominant haplotype H1 found in Ireland in this study. Interestingly, the same study found no variation at six chloroplast microsatellite loci, in comparison with the 23 haplotypes from the six loci used in the present study. This is most likely due to the fact that these six loci are "universal", rather than species-specific, and tend to exhibit limited polymorphism in general (Provan et al. 2001).

The level of population differentiation observed for nuclear markers ($\Phi_{ST(n)}$ = 0.0092) is over an order of magnitude lower than those reported previously for *C. monogyna* and other *Crataegus* species. Ferrazzini et al. (2008) reported a value of $\Phi_{ST(n)}$ = 0.203 in six populations of *C. monogyna* from northern Italy, despite the study area being approximately the same as that in the present study. This discrepancy is almost certainly due to the different properties of the markers used, since it has been shown previously that measures of population differentiation based on RAPD phenotypes, as estimated by Ferrazzini et al. (2008), are generally around one order of magnitude greater than those estimated using multi-allelic codominant markers (Isabel et al. 1999). Microsatellite analyses of natural populations of *C. douglasii* and *C. suksdorfii* yielded $\Phi_{ST(n)}$ values of 0.22 and 0.37 respectively, although the latter fell to 0.15 when only diploid samples were included, as much of the differentiation

was postulated to arise from some degree of reproductive isolation between individuals with differing ploidy levels (Lo et al. 2009). It was also inferred that gene flow was further reduced between populations as a result of apomixis.

The fourfold higher Φ_{ST} value observed for chloroplast markers ($\Phi_{ST(c)} = 0.0385$) compared to that for nuclear markers ($\Phi_{ST(n)} = 0.0092$) is consistent with the fact that chloroplast markers have a lower effective population size, and are maternally inherited in angiosperms and thus dispersed via seed. As a result, they generally show higher levels of genetic structuring (Provan et al. 2001). Nevertheless, both values are very low, indicating extensive gene flow across the study area via both pollen and seed. The ratio of pollen:seed gene flow ($r = 2.515$) is among the lowest reported for a range of plant species (Ennos 1994), indicating efficient seed dispersal. In hawthorn, this is primarily via ingestion by mammals and birds, and the latter, particularly thrushes (*Turdus* sp.), are most likely responsible for long-distance dispersal events (García and Chacoff 2007). Extensive pollen flow is also expected due to the wide foraging ranges of the bees and flies that pollinate the species (Steffan-Dewenter et al. 2002).

The BAPS analysis indicated that all hawthorn populations studied, including that from Coolure in the Republic of Ireland, belonged to a single genetic cluster. This was also the case in ash (Beatty et al. 2015a), and similar to that in alder, where 25 of the 26 populations studied were assigned to the same cluster (Beatty et al. 2105b). Thus, as in the previous studies, there is no obvious genetic structuring that could be used as the basis for management units or seed zones. Nevertheless, examination of the chloroplast haplotypes, which are maternally inherited and thus indicative of seed diversity, can further inform management decisions with respect to replanting and restocking. Populations like that at Portaferry, which contains not only seven of the eleven non-unique haplotypes but also four unique haplotypes, and that at Eglinton, which has the highest overall chloroplast diversity,

would appear to represent good sources of genetic variation. In addition, the Glenarm population is the only one which is not dominated by the most common haplotype (H1), and thus may merit special consideration. Conversely, the Belleek and Drumshanbo Lough populations have the least amount of haplotypic diversity, along with significant F_{IS} values, and thus may not represent the best options for the acquisition of seeds for restocking. The lack of differentiation between the Northern Ireland populations and Coolure might indicate that, on a broader scale, seeds could be sourced from anywhere in Ireland, but further work on populations from the Republic of Ireland would be needed to confirm this, and to identify any potential structuring of chloroplast variation. It should be noted that the use of ostensibly neutral markers, as in the present study, may not reflect local adaptive variation, which would only be evident from reciprocal transplant or common garden experiments.

It is concluded that extensive pollen- and seed-mediated gene flow occurs in the populations of hawthorn analysed in the present study, giving rise to high levels of genetic diversity but low levels of genetic differentiation between populations. These results mirror those from other woodland tree species from the same region. Thus, there may be an emerging picture that management of these species may be most efficiently carried out at a regional level, although a more holistic approach might also target rare or high chloroplast (seed-specific) variation. Such approaches might not be so suitable for other common tree species such as hazel (*Corylus avellana*), which have very large seeds and thus may have much more limited dispersal potential. Further information on such species is needed to complement the present and earlier findings.

Acknowledgements

James Brown and Caroline Finlay's PhD studentships were funded by the Department for Agriculture and Rural Development, Northern Ireland. We are also indebted to the Northern Ireland Environment Agency for additional funding, to landowners for allowing access to their property, and to Dr Sally Montgomery, Marion Chapalain and Thomas Finlay for assistance with sampling. We are grateful to two Reviewers for helpful comments on the original version of the manuscript.

Data archiving statement

All data will be deposited in DRYAD on acceptance.

281 **References**

282

283 Bacles CFE, Ennos RA (2008) Paternity analysis of pollen-mediated gene flow for *Fraxinus*
284 *excelsior* L. in a chronically fragmented landscape. Heredity 101: 368–380.
285 Beatty GE, Brown JA, Cassidy EM, Finlay CMV, McKendrick L, Mongomery WI, Reid N,
286 Tosh DG, Provan J (2015a) Lack of genetic structure and evidence for long-distance
287 dispersal in ash (*Fraxinus excelsior*) populations under threat from an emergent fungal
288 pathogen: implications for restorative planting. Tree Genet Genomes 11:53
289 Beatty GE, Montgomery WI, Tosh DG, Provan J (2015b) Genetic provenance and best
290 practice woodland management: a case study in native alder (*Alnus glutinosa*). Tree
291 Genet Genomes 11, 92.
292 Beatty GE, Montgomery WI, Spaans F, Tosh DG, Provan J (2016) Pure species within a
293 continuum of genetic and morphological variation: sympatric oaks at the edge of their
294 range. Annals Bot 117:541-549.
295 Burel F, Baudry J (1995) Social, aesthetic and ecological aspects of hedgerows in rural
296 landscapes as a framework for greenways. Landscape Urban Plann 33:327–340.
297 Campbell CS, Greene CW, Dickinson TA (1991) Reproductive biology in subfam. Maloideae
298 (Rosaceae). Syst Botany 16:333–349.
299 Christensen KI (1992) Revision of *Crataegus* sect. *Crataegus* and nothosect. *Crataeguineae*
300 (Rosaceae-Maloideae) in the Old World. Syst Botany Monographs 35:1–199.
301 Clapham AR, Tutin TG, Moore DM (1990) *Flora of the British Isles*. Cambridge University
302 Press, Cambridge, UK.
303 Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation
304 between populations. Genetics 163:367–374.

305 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf
306 tissue. Phytochem Bull 19:11–15.
307 Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant
308 populations. Heredity 72:250-259.
309 Ennos RA, Worrell R, Malcolm DC (1998) The genetic management of native species in
310 Scotland. Forestry 71:1–23.
311 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from
312 metric distances among DNA haplotypes - application to human mitochondrial DNA
313 restriction data. Genetics 131:479-491.
314 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform
315 population genetics analyses under Linux and Windows. Mol Ecol Resources 10:564–
316 567.
317 Ferrazzini D, Monteleone I, Belletti P (2008) Small-scale genetic diversity in oneseed
318 hawthorn (*Crataegus monogyna* Jacq.). Eur J Forest Res 127:407–414.
319 Fineschi S, Salvini D, Turchini D, Pastorelli R, Vendramin GG (2005) *Crataegus monogyna*
320 Jacq. and *C. laevigata* (Poir.) DC. (Rosaceae, Maloideae) display low level of genetic
321 diversity assessed by chloroplast markers. Plant Syst Evol 250:187–196.
322 García D, Chacoff NP (2007) Scale-dependent effects of habitat fragmentation on hawthorn
323 pollination, frugivory, and seed predation. Conserv Biol 21:400-411.
324 Gosler AG (1990) Introgressive hybridization between *Crataegus monogyna* Jacq. and
325 *Crataegus laevigata* Poiret DC. in the Upper Thames Valley, England, UK. Watsonia
326 18:49–62.
327 Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices
328 (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

329 Gyan KY, Woodell SRJ (1987) Analysis of insect pollen loads and pollination efficiency of
330 some common insect visitors of four species of woody Rosaceae. Functional Ecology
331 1:269-274.
332 Herbert R, Samuel S, Pattison G (1999) *Using Local Stock for Planting Native Trees and*
333 *Shrubs.* Forestry Commission Practice Note 8. Forestry Commission, Edinburgh, UK.
334 Howe HE, Smallwood J (1982) Ecology of seed dispersal. Ann Rev Ecol Systematics
335 13:201–228.
336 Isabel N, Beaulieu J, Thériault P, Bosquet J (1999) Direct evidence for biased gene diversity
337 estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. Mol
338 Ecol 8:477-483.
339 Jacobs JH, Clark SJ, Denholm I, Goulson D, Stoate C, Osborne JL (2009) Pollination biology
340 of fruit-bearing hedgerow plants and the role of flower-visiting insects in fruit-set. Annals
341 Bot 104:1397–404.
342 Levine JM, Murrell DJ (2003) The community level consequences of seed dispersal patterns.
343 Ann Rev Ecol Evol Systematics 34:549–574.
344 Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, van de Weg E, Gessler C
345 (2002) Development and characterisation of 140 new microsatellites in apple (*Malus* x
346 *domestica* Borkh.). Mol Breeding 10:217–241.
347 Lo EYY, Stefanović S, Dickinson TA (2009) Population genetic structure of diploid sexual
348 and polyploid apomictic hawthorns (*Crataegus*; Rosaceae) in the Pacific Northwest. Mol
349 Ecol 18:1145–1160.
350 Müller-Starck G, Baradat P, Bergmann F (1992) Genetic variation within European tree
351 species. New Forests 6:23–47.
352 Nathan R, Muller-Landau HC (2000) Spatial patterns of seed dispersal, their determinants
353 and consequences for recruitment. Trends Ecol Evol 15:278–285.

354 Pollard E, Hooper MD, Moore NW (1974) *Hedges*. Collins Press, London, UK.
355 Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence
356 repeats. Trends Plant Sci 1:215-222.
357 Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for
358 studies in plant ecology and evolution. Trends Ecol Evol 16:142–147.
359 Provan J, Beatty GE, Hunter AM, McDonald RA, McLaughlin E, Preston SJ, Wilson S
360 (2008) Restricted gene flow in fragmented populations of a wind-pollinated tree. Conserv
361 Genet 9:1521-1532.
362 Rackham O (2008) Ancient woodlands: Modern threats. New Phytologist 180:571–586.
363 Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for
364 exact tests and ecumenicism. J Hered 86:248–249.
365 Schroth MN, Thomson SV, Hildebrand DC, Moller WJ (1974) Epidemiology and control of
366 fire blight. Ann Rev Phytopathol 12:389-412.
367 Steffan-Dewenter I, Müntzenberg U, Bürger C, Thies C, Tscharntke T (2002) Scale-
368 dependent effects of landscape context on three pollinator guilds. Ecology 83:1421–1432.
369 Stowell EA, Backus MP (1966) Morphology and cytology of *Diplocarpon maculatum* on
370 *Crataegus* L. I. The *Entomosporium* stage. Mycologia 58:949-960.
371 Sutherland BG, Belaj A, Nier S, Cottrell JE, Vaughan SP, Hubert J, Russell K (2010)
372 Molecular biodiversity and population structure in common ash (*Fraxinus excelsior* L.) in
373 Britain: implications for conservation. Mol Ecol 19:2196-2211.
374 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
375 structure. Evolution 38:1358–1370.

Table 1 Details of hawthorn (*Crataegus monogyna*) populations studied. Type – H hedgerow, W woodland; Lat/Long given in degrees; *N* – number of individuals analysed; A_R – allelic richness; H_O – observed heterozygosity; H_E – expected heterozygosity; F_{IS} – inbreeding coefficient (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS non-significant); H1-H11 – frequency of chloroplast haplotypes (Un – unique haplotypes found in a single individual); *H* – gene diversity.

No	*Name*	*Type*	*Lat*	*Long*	*Nuclear*					*Chloroplast*													
			(N)	*(W)*	N	A_R	H_O	H_E	F_{IS}	N	*H1*	*H2*	*H3*	*H4*	*H5*	*H6*	*H7*	*H8*	*H9*	*H10*	*H11*	*Un*	H
1	Portaferry	H	54.394	5.527	28	12.074	0.783	0.838	0.067**	30	17	2	2	2	-	1	-	1	1	-	-	4	0.681
2	Glenarm	W	54.951	5.965	25	8.779	0.729	0.772	0.058*	25	10	2	12	-	-	-	1	-	-	-	-	-	0.627
3	Kilkeel	H	54.065	6.087	30	10.231	0.747	0.809	0.077*	30	22	2	1	3	2	-	-	-	-	-	-	-	0.458
4	Rea's Wood	W	54.710	6.234	30	10.231	0.782	0.802	0.026^{NS}	30	21	4	4	1	-	-	-	-	-	-	-	-	0.490
5	Breen Wood	W	55.140	6.240	29	11.056	0.714	0.804	0.114***	30	20	1	1	5	2	-	-	1	-	-	-	-	0.538
6	Loughbrickland	H	54.292	6.322	30	10.188	0.763	0.798	0.044^{NS}	30	25	2	-	-	1	-	1	-	-	1	-	-	0.308
7	Ballyronan	H	54.701	6.553	27	10.288	0.778	0.793	0.020^{NS}	30	26	-	2	1	1	-	-	-	-	-	-	-	0.251
8	Kilrea	H	54.943	6.573	30	10.273	0.795	0.801	0.008^{NS}	30	25	3	2	-	-	-	-	-	-	-	-	-	0.301
9	Portrush	H	55.195	6.577	30	10.356	0.731	0.795	0.082**	30	21	4	2	3	-	-	-	-	-	-	-	-	0.494
10	Peatlands Park	W	54.486	6.617	29	11.798	0.799	0.812	0.016^{NS}	30	20	2	5	1	-	-	-	-	-	-	-	2	0.538
11	Aghadowey	H	55.015	6.645	30	10.637	0.789	0.812	0.029^{NS}	29	23	1	2	2	-	-	-	-	-	-	1	-	0.372
12	Errigal Glen	W	54.975	6.729	30	11.317	0.775	0.821	0.057*	30	25	3	-	1	1	-	-	-	-	-	-	-	0.303
13	Keady	H	54.282	6.735	30	10.860	0.836	0.818	-0.022^{NS}	30	22	2	2	2	1	-	-	-	-	-	1	-	0.462
14	Killylea	H	54.373	6.841	22	11.001	0.815	0.824	0.011^{NS}	30	24	2	1	1	-	1	-	-	-	1	-	-	0.363
15	Bannagher Glen	W	54.885	6.957	30	9.393	0.739	0.782	0.056*	30	26	-	-	-	-	-	3	-	-	-	-	-	0.246
16	Eglinton	H	55.047	7.214	28	11.612	0.753	0.834	0.099***	30	16	2	2	2	-	-	-	1	-	-	-	4	0.713
17	Plumbridge	H	54.743	7.227	27	10.464	0.806	0.808	0.003^{NS}	28	18	4	1	5	-	-	-	-	-	-	-	-	0.553
18	Lisnaskea	H	54.240	7.395	19	9.649	0.768	0.817	0.061^{NS}	20	14	2	1	-	3	3	-	-	-	-	-	-	0.500
19	Crom	W	54.167	7.454	30	10.271	0.748	0.787	0.050*	30	24	2	2	1	-	-	-	-	-	-	-	1	0.361
20	Drumshanbo Lough	H	54.654	7.493	27	10.580	0.709	0.775	0.087**	30	27	1	-	1	-	-	-	-	1	-	-	-	0.193
21	Lough Erne	H	54.454	7.714	27	10.259	0.723	0.781	0.076**	30	23	3	-	2	1	1	-	-	-	-	-	1	0.409
22	Marble Arch	W	54.266	7.814	30	10.834	0.744	0.808	0.080**	30	26	1	-	1	2	2	-	-	-	-	-	-	0.251
23	Belleek	H	54.465	8.107	29	10.008	0.684	0.769	0.122***	30	28	1	1	-	-	-	-	-	-	-	-	-	0.193
24	Coolure	W	53.671	7.360	30	11.185	0.789	0.814	0.031^{NS}	29	21	3	-	5	-	-	-	-	-	-	-	-	0.451

Table 2 Hawthorn (*Crataegus monogyna*) nuclear microsatellite loci analyzed in this study. Loci were originally described in Liebhard et al. (2002). Cycles – number of cycles used in PCR; N – number of alleles; A_R – mean allelic richness; H_O – mean observed heterozygosity; H_E – mean expected heterozygosity; F_{IS} – mean inbreeding coefficient

Locus	Cycles	N	Range (bp)	A_R	H_O	H_E	F_{IS}
CH01F02	45	36	213-305	15.600	0.924	0.927	0.004
CH02D11	40	31	159-221	12.067	0.787	0.768	-0.030
CH03C02	40	33	117-193	14.394	0.760	0.871	0.130
CH04G04	40	16	171-213	6.433	0.767	0.768	0.001
CH05D04	40	33	188-258	16.333	0.846	0.931	0.094
CH05D11	45	19	183-231	7.182	0.654	0.709	0.095
CH05G07	40	26	174-224	12.689	0.771	0.883	0.116
CH05G11	30	12	215-247	4.220	0.600	0.568	-0.033

Table 3 Hawthorn (*Crataegus monogyna*) chloroplast microsatellite loci analyzed in this study.

Multiplex	Accession	Region	Repeat	Primers*	Alleles (bp)
1	EU500411	*rpl*2-*trn*H intergenic	$(A)_9$	ATAAAAACAAAAATAGGAGTAATTAATTGTGAC	100,101
				TTCTTAATAAATGATTTGCTACAAAAGG	
	JQ390913	*rpl*20-*rps*12 intergenic	$(T)_{13}$	TATAACCTTCCCGACCACGA	119,121,122,123
				ATTTACTACTTTTATGTGTTTTTGATACCT	
	JQ392044	*trn*G-*trn*S intergenic	$(A)_9$	GATTCGTTGGAACAATAAATGG	135
				GGATTGAAAGAGCCCTTCATAA	
2	JQ391567	*rpl*16 intron	$(T)_9$	TTGCTTTACAACCCATAATCAGA	159,161,162,163,164,166
				ACCAACTCATCACTTCGTGTT	
	HG764984	*trn*H-*psb*A intergenic	$(T)_{13}…(A)_{10}$	AGATAAAATACAACCTAAATTGAAAACTT	195,196,219,220,221,222,223
				ATATATGAGTTCTTGAAAGTAAAGGAGTAA	
	FJ395291	*atp*F-*atp*H intergenic	$(T)_{10}$	CATTTTTCATATGATATCCTCTTTCTT	244,246
				CGGGTACCTAATTCTAATAAGTATCATTC	

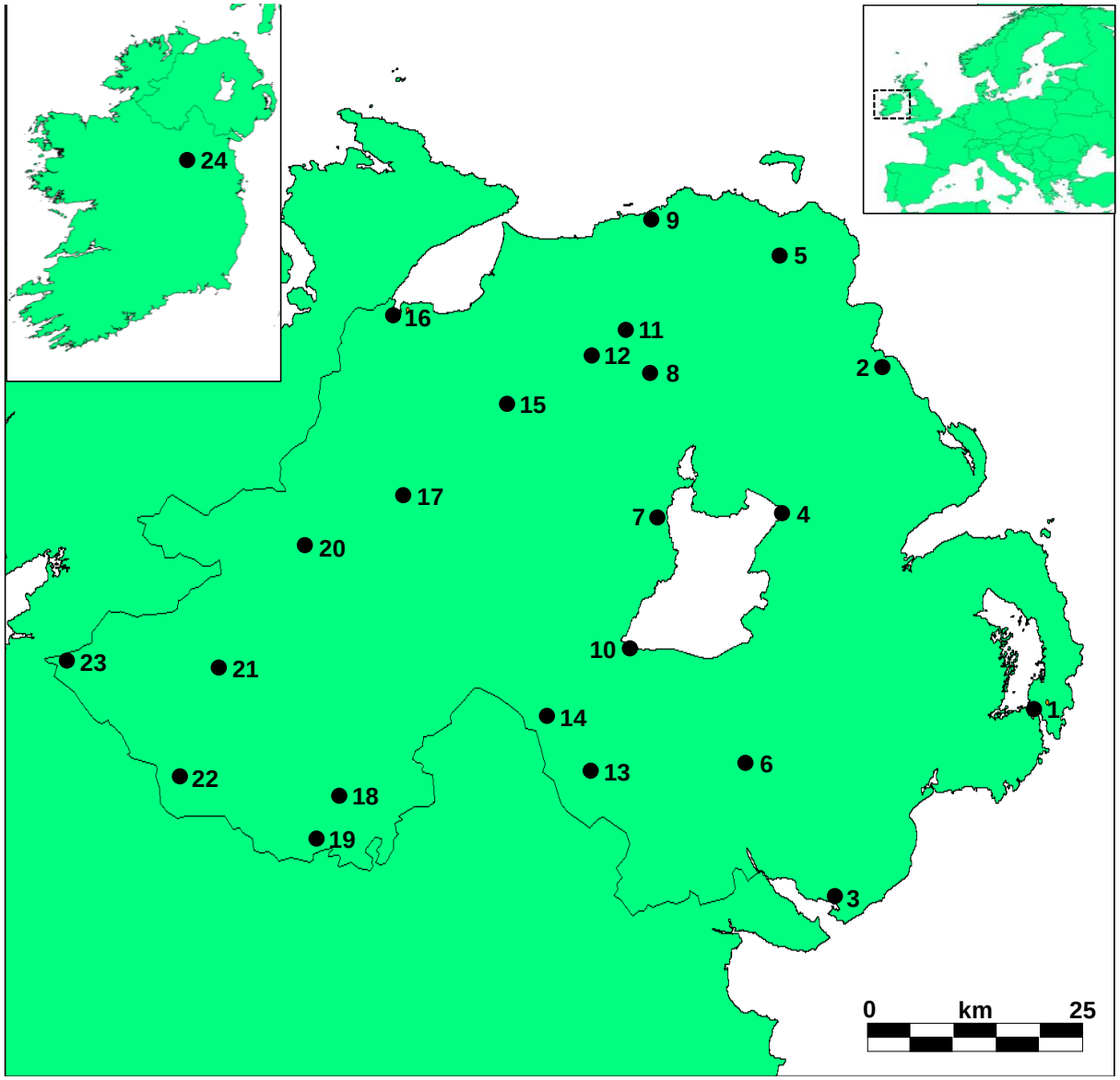
* Forward tailed with CACGACGTTGTAAAACGAC; Reverse tailed with GTGTCTT

Table 4 Analysis of molecular variance (AMOVA) for nuclear and chloroplast microsatellite markers analysed in hawthorn (*Crataegus monogyna*).

Markers	Source of variation	Sum of squares	Variance	% variation
Nuclear	Among populations	105.747	0.02811	0.92
	Within populations	4006.636	3.01251	99.08
Chloroplast	Among populations	10.378	0.00833	3.85
	Within populations	140.752	0.2079	96.15

Figure Legend

Fig. 1 Locations of sites sampled in this study. Numbers correspond to those in Table 1. The dashed box in the top-right map show the area in Europe which the study was carried out.


24
9
5
16
11
12
8
2
15
17
7
4
20
10
23
21
14
1
22
13
6
18
19
3
0
km
25